Measuring hardness of problems in evolutionary computations using Markov-processes

Márk Jelasity, Boglárka Tóth and Tamás Vinkó

In the recent years it has been clearified that evolutionary computations (EC) works very effective on real-life problems. However, the theory of EC is far from clear, finding a good characterization of problem difficulty is one of the main lines of EC theory research. So the question is, which problems are hard/easy to solve with the algorithms of EC. There are many suggested measure, for example epistasis variance and fitness-distance correlation (FDC). Epistasis variance is not so reliable and not so easy to compute. The FDC is much better, but there are some counterexample, which missleading this measure too. In this paper a new measure is suggested based on statistical properties of trajectories. These properties are approximated with the help of a heuristic based on transition probabilities between the elements of the search space, which derived from the Markov-processes. The basic idea is to examine the trajectories of the space with respect to a given operator and stopping criterion. The ending points of these trajectories form a very interesting set: these are the points the search is expected to converge. Based on the transition probabilities it is also possible to approximate the expected number of evaluations needed to get from a point to a given other point. This values can be used as distance measures and plots can be drawn that depict the convergence relations. With these definitions it is also possible to introduce a deceptiveness coefficient: a number from $[0, 1]$, which characterizes the problems: 0 indicates that the problem is misleading, 1 means that it is very friendly and there are transitions for other cases.

With our method any mutation or other genetic operator can be used, however only the mutation operator examined because the probabilities mentioned above need a huge amount of calculation. So this paper deal with relative small problems, but we don't model the fitness function explicitly, it is handled as a black box. Though the computational efforts are higher, the method is still feasible due to some techniques that speed up the convergence.

The layout of our paper is the following: after the introduction, in the section 2 some definitions and notions are given. In section 3 our method is demonstrated on four well-known problems from EC literature: ridge function, long path problem, a fully deceptive function and a combinatiorial problem: the subset sum problem. For the problems, three types of figures are introduced: iteration, deceptiveness and endpoint figures. In this section some explanation is given on how to read these figures. Finally our method is validated via some empirical results.